

# An Information Model for Medical Events

Daniel J. Essin, MD LAC+USC Medical Center

Thomas L. Lincoln, MD LAC+USC Medical Center and RAND Corporation

## ABSTRACT

Information gathered during the healthcare process is lost when forced into rigidly structured record-oriented databases. By contrast, content can be difficult to manipulate if stored as unstructured text. Spurred by the requirements of electronic publishing, military procurement and the Internet, new robust standards for structuring documents have been developed and deployed. These standards can provide a foundation for a document-based Electronic Medical Record System. In order to fully exploit this added flexibility, an information model is necessary to define both the direct and contextual content of documents. Once context, as well as fact, are recorded in formal structures, inferential techniques can either selectively extract knowledge and data from documents or aggregate data to create summaries so that all interested and authorized parties have a better chance of meeting their information needs from a single, permanent data source.

## BACKGROUND

Information gathering is a cognitive activity. Our environment is filled with objects and situations that have the potential to yield information. The information that comes to our attention does so either as the direct result of our experiences, or as the result of reflecting on those experiences [1]. The discovery or recognition of these occurrences, and the subsequent generation of information from them, takes place in a discrete and episodic fashion rather than continuously. An information model describes when this cognitive activity will be translated into a machine processable representation and what content will be included.

Individuals intuitively know when to vary the structure of their notes to effectively capture or represent relevant context. To represent this informational environment in an electronic format, structure and content need be modeled separately so that the relationship between them can be

dynamically adjusted to meet the evolving needs and expectations of the human users.

Models are abstract representations of concepts, forgoing detail in order to isolate and clarify a set of core properties. The design and function of any complex system depends on the participation of many concepts which, when expressed along with their inter-relationships, constitute the system's architecture (or model.) The architecture of Electronic Medical Records Systems (EMRS) can be expected to include models of requirements, workflow, transactions, security, data structure, denotational semantics (the semantics that provide the symbols and names necessary to signify meaning), and information content.

Of these models, perhaps the most fundamental are 1) the requirements model that describes what the system is expected to do and 2) the information model that identifies the information content that must be captured to fulfill the requirements. The popular system development methodologies contain implicit models of structure and denotational semantics. Although these can interfere with the full realization of requirements or with the graceful evolution of an implementation the "best" structure or semantic model cannot compensate for the failure to capture relevant information. Many successful healthcare information systems have responded to these constraints by restricting themselves to the processing of highly structured documents whose content can be mapped onto formal database structures.

Loosely structured documents offer a unique alternative that can address a broader range of applications and information [2,3]. There is a growing body of standards that address the formal structure and processing of loosely structured documents. For example, HyTime<sup>\*</sup> [4], built on SGML<sup>†</sup> [5] provides a notation that defines a generalized hierarchy of document

---

<sup>\*</sup> Hypermedia/Time-based structuring language ISO/IEC 10744

<sup>†</sup> Standard Generalized Markup Language ISO/IEC 8879

types and provides a means of recording and processing them. Since loosely structured documents can be formally described and processed, they are a good candidate for the foundation upon which to build an EMRS.

## THE INFORMATION MODEL

The objective of the information model presented here is to address six observations. 1) Facts originate in events. 2) Facts require context to be informative. 3) Context is always rich and highly variable. 4) The temporal association of facts and events is a significant component of context. 5) Context is difficult to reconstruct; if it is not captured "up-front" it is usually lost forever. 6) Although information can be modeled separately, in practice the underlying structure and semantic models have a strong influence on the ability to fulfill the demands of the information model.

The moments during which information is recognized or synthesized are called *Information Laden Events* (ILE). If an individual makes only a "mental note" of information, it will have no direct impact on the actions of others now or in the future. The act of recording an ILE instantiates the factual or informational content of the event as part of the universe of known data. The term *Information Bearing Object* (IBO) [6]. will be used to describe the records of those events which become part of the accessible information (knowledge) base by virtue of being recorded by people or by devices or processes set into motion by people. The model describes the components that must (or may) be present within an IBO to insure that the facts and the context of the event are captured in a machine processable representation.

Facts require context in order to be considered information. The model therefore requires a statement of "WHAT" happened (including the EVENT TYPE) to provide the fact as well as statements of "WHO", "WHERE", "WHEN", "WHY" and "HOW" to provide the context. Different event types have different minimal requirement for context below which the IBO would not be informative. Beyond that minimum, there are a wide variety of additional context-defining items that could be supplied. The information content increases as more context is recorded. The importance or relevance of additional information

depends on the nature of the primary event and what unique circumstances may have occurred.

The following is a non-exhaustive list of the informational components that may be necessary to fully describe the content and context of an ILE. Each component can be expected to have substructure depending upon the nature of the event to be documented and the type of document being used to record the event.

WHO (acting in what capacity or role)

To whom did it happen?

A person, a group or association of people

A legal entity, an *ad hoc* association

Who did it?

Who recorded it, checked it, approved it?

Who looked at it, changed it, etc.?

WHAT kind of event is it?

WHAT thing(s) were involved?

WHAT detailed information was collected?

Data elements ("fields")

Structured notes

Narrative

WHAT special circumstances were/will be involved?

WHAT other events are linked/related to this event?

WHAT happens next?

WHAT version is this?

WHERE did it/will it happen?

WHEN

When did it/will it happen?

When did it/will it expire/be retired?

HOW did it/will it happen?

Orders

Simple, Complex

Procedures

WHY did it/will it happen?

Assessment, Commentary, Annotation

To continue this formulation it is necessary to establish the fundamental properties of each component of the model and to explore the denotational semantics that may be required to represent them. They are:

1) persons

2) objects

a) with a unique physical presence other than persons e.g. items of capital equipment with serial numbers, etc.,

- b) objects that have a non-unique physical presence
  - such as pi-mesons, Band-Aids, individual fresh carrots, etc.
- 3) points on a timeline
- 4) points in space
- 5) Assertions or retractions of fact or opinion along with supporting context. There may be many types of events including:
  - a) informational (descriptive) events
  - b) identity-defining events
  - c) complex entity, organization and association describing events
  - d) role-defining events
  - e) knowledge-defining events
  - f) rule-defining events
  - g) opinion-defining events

## Persons

"Person" references literally represent people, not an abstract class. There are no subtypes of people. With the exception of the genetically determined links (if known) to mother and father, relationships between people are determined by applying inference to a collection of identity-defining and relationship-defining IBO's using a collection of rules that define relatedness. Several examples will illustrate the intended interactions.

The determination of siblings and family groupings requires inferential queries. To find an individual's brothers requires retrieving a rule (that defines a brother as any male offspring of an individual's mother), then searching for birth events involving the UID of the mother that produced male children followed by the retrieval of the necessary IBO's.

Age is another datum that must be inferred since it is not constant and can have a variety of context-dependent values. Even with birthdate information available, the age cannot be computed until the age context is defined. Is it the age today or the age at the time of some event?

Something as apparently simple as a person's name is subject to similar ambiguity. Just as with age, the question must be asked - Name as of "When" and for "What" reason? In many settings the name recorded as part of the birth event is not the legal name since there are a number of circumstances in which names change such as adoption, marriage, etc.

## Inanimate Objects

Inanimate objects fall into two categories. Some objects are significant because they are expensive, dangerous or must have an individual identity for some reason. Such objects have some properties in common with persons and other complex entities. Other inanimate objects, especially consumable supplies, are treated differently in purchasing, receiving, inventory, distribution, and usage accounting.

Certain IBO's, such as documentation of surgical procedures, will need to indicate that various equipment and supplies were brought into the operating room, which were used, opened, applied to the patient, disposed of, etc. A semantic network describing the similarity between various objects would facilitate fuzzy searches such as "Find everyone who had intravenous therapy on March 12".

## Spatial and Temporal Location

Every IBO should contain at least the date, time and location when it was created. In order for IBO's to be meaningful in the future or if transmitted to distant locations, the data that place an event in time and space must be global in nature so that the meaning is not lost. Most existing systems satisfy this requirement for dates but not for time values. Times should be stored as Greenwich Mean Time (GMT) and all locations should be stored as a latitude, longitude, and elevation allowing for the indirect determination of the GMT offset.

The practice of describing locations in relative or colloquial terms is a great source of ambiguity as functions are moved from place to place while retaining familiar names. The GPS (Global Positioning System) Coordinates of any location can now be readily determined and make it possible to describe locations in absolute terms. Geographic Information Systems can translate the coordinate data into more familiar terms as needed. In addition, date and location data make it possible to properly interpret time both relative to Daylight Savings Time and changes in time zone.

## Roles

Persons may assume, or be assigned, a large and unpredictable number of roles. The set of allowable

roles changes as the organization evolves. Using IBO's to create and assign roles eliminates the need to identify the universe of allowable roles before a system is actually built. Many development methodologies, make role information a structural component of system design that is difficult to augment or eliminate at a later date.

Complex entities, represent a special case of role-definition. They originate in events that assert their creation. Subsequent events can then assign persons or other complex entities to roles within the new organization or association. Examples of complex entities are corporations, committees, ward-teams, group practice associations, etc. While in existence, complex entities can assume many, but not all, of the roles that can be assumed by persons.

### Information and Knowledge

Several other types of knowledge are required to integrate a complete system:

- 1) Domain-specific knowledge, e.g. medical terminology and medical diagnoses;
- 2) Application structure and control information;
- 3) Ordered or unordered links between events;
- 4) Application defined functions and rules;
- 5) Features that will support activity-tracking;
- 6) Rules defining security and user permissions and capabilities;
- 7) and, the ability to include multimedia data as part of "What", "How", and "Why" notations.

Denotational semantics are the subject of much active research. Organizations will have different needs in this area if they are centers of excellence in highly specialized domains or have evolved a unique institutional culture. The purpose of the information model is to explicitly accommodate a variety of denotational semantics, not to impose particular choices. When organizations using different semantics need to exchange information, pragmatic translations between semantics can be negotiated providing that a consistent information model is in use.

Knowledge can be represented using the same structures that represent "data." This will allow knowledge and data to be managed by the same tools. Likewise, the rules that define what data mean and the various ways that data should be validated, displayed and in-

terpreted should reside in the database as computable entities.

### Simple vs. Compound IBO's

Some events that occur take a long time to complete. The need to keep contemporaneous documentation will mean that some events may be composed of multiple IBO's. These can be termed Compound Events. The posting of correcting entries and other types of audit information may convert an event from simple to compound. Thus it will be necessary to provide retrieval functions that will aggregate the components of a compound event and merge them into a single record for display purposes.

### Formal Elements of the Model

IBO's are composed of a series of statements about context and information. Both the IBO's and the statements within them are produced by generating functions.

Abstractly we have:

$\mathcal{D} = \{D_1, D_2, \dots, D_n\}$  The set of domains of denotational semantics.

Every system uses a variety of denotational semantics. Some may adopted from external sources such as English, SNOMED [7], the MIB Semantic Model [8] or Univ. of Utah's Event Definitions [9] while others may be developed locally such as a hospital formulary or a list of approved abbreviations.

$\mathcal{E} = \{ \text{person, unique inanimate object, nonunique inanimate object, association} \}$  The set of entity types.

$E = \{ (\epsilon, \text{role}) \}$  The set of entity references as entity-role tuples.

Entities are an important part of most statements since they are either involved as the subjects or objects of actions. References to entities usually include a role qualifier such as "a person acting in a capacity" or "a thing used for a purpose."

$\mathcal{C} = \{ \text{event\_type, location, time, etc.} \} \forall \mathcal{C}, \mathcal{C} \in \mathcal{D}$ . The set of context-describing elements mapped to various domains of denotational semantics.

$\mathcal{K} = \{ \text{data, knowledge, opinion, rules, etc.} \} \forall \mathcal{K}, \mathcal{K} \in \mathcal{D}$   
The set of knowledge-describing elements.

$A = \{ A_1, A_2, \dots, A_n \} \forall A, A \in \delta$ . The set of actions and processes.

$\rho = \{ \text{permissions, capabilities} \} \forall \rho, \rho \in \delta$ . The set of security-describing elements.

Each of the elements used to describe context, information/knowledge, actions and permissions is described using a specific denotational method.

$\delta = \{ S_1, S_2, \dots, S_n \}$  The set of statement constructor functions.

$\theta = \{ O_1, O_2, \dots, O_n \}$  The set of IBO constructor functions.

$X = \delta(E, C, A, \rho)$  A contextual statement generated by statement construction function

$I = \delta(E, C, A, \rho, \kappa)$  An informational statement generated by statement construction function.

$IBO = \theta(\{ X_1, X_2, \dots, X_n \}, \{ I_1, I_2, \dots, I_n \})$  an IBO is composed of a group of context statements followed by a group of informational statements.

Statements and IBO's are the output of a set of functions that take entity references, context, knowledge, actions and permissions as inputs. The functions can take many forms depending on the tactics and the strategy of a particular implementation. The components of these functions represent a convenient of checklist to be satisfied by any implementation.

## SUMMARY

Regardless of the techniques used to structure documents, an Information Model is necessary in order to organize and describe the content. The information model delineates the content that must be present to insure that a description of an event is fully meaningful. The elements of this model can be applied to object-oriented structures or relational constructs as well as to loosely structure documents. Up to now, the hierarchical properties of objects and the notational complexities of relational databases introduced scale-up problems. These difficulties are markedly accentuated by the need to complete the design before releasing a product. The most common response of developers has been to avoid the added complexity that would be needed to fully qualify the context of an event. Requiring adoption of and adherence to an explicit Information Model is one way to insure that the next generation of systems are capable of capturing the context necessary to make the data informative over a long period of time.

## REFERENCES

1. Norman D: "Things That Make Us Smart: Defending human attributes in the age of the machine," Addison-Wesley, 1993.
2. Lincoln TL, Essin DJ and Ware WH: The Electronic Medical Record: A Challenge for Computer Science to Develop Clinically and Socially Relevant Computer Systems to Coordinate Information for Patient Care and Analysis. The Information Society Journal, Vol. 9, No. 2 (Apr-Jun 1993).
3. Essin DJ: Intelligent Processing of Loosely Structured Documents as a Strategy for Organizing Electronic Health Care Records. Methods of Information in Medicine, 32:265-68, 1993.
4. Newcomb S, Kipp N, and Newcomb V: The 'HyTime' Hypermedia/Time-based Document Structuring Language, Commun. ACM, 24:11, (November 1991), 67-83.
5. Goldfarb C: "The SGML Handbook," Oxford University Press, 1990.
6. Guha RV and Lenat DB: Enabling Agents to Work Together, Commun. ACM, 37:7, (July 1994), 127-142.
7. SNOMED International: the systematized nomenclature of human and veterinary medicine [3rd ed.], Editors: Roger A. Cote, et al., College of American Pathologists & American Veterinary Medical Association, Schaumburg, IL, 1993..
8. Gottschalk HW and Wittenber J: "MIB Software - Semantic Model: ASN.1 Messaging for a pulse Oximeter," 14th Annual Symposium on Computer Applications in Medical Care, IEEE Computer Society Press, (November 1990), 220-225.
9. Fu LS, Bouhaddou O, Huff, SM, Sorenson DK and Warner HR: "Toward a Public Domain UMLS Patient Database," 14th Annual Symposium on Computer Applications in Medical Care, IEEE Computer Society Press, (November 1990), 170-174.